

DOCTRINA

El derecho al control humano: Una respuesta jurídica a la inteligencia artificial

The right to human control: A legal response to artificial intelligence

Carolina Sánchez Vásquez  y José Alberto Toro-Valencia 

Universidad EAFIT, Colombia

RESUMEN La inteligencia artificial es uno de los mayores desarrollos tecnológicos del siglo XXI y tiene la capacidad para incidir de manera positiva, pero también negativa en la vida del ser humano. Así, el uso de sistemas de inteligencia artificial puede vulnerar derechos como la igualdad, la privacidad, el debido proceso y la libertad de expresión. En ese escenario, el derecho necesita responder a los nuevos retos que se plantean y ofrecer respuestas jurídicas vanguardistas. Una de ellas es el reconocimiento del control humano como un nuevo derecho para garantizar la veeduría humana en el diseño y desarrollo de dichos sistemas con el fin de evitar afectaciones a otros derechos ya existentes.

PALABRAS CLAVE Inteligencia artificial, derechos humanos, control humano.

ABSTRACT Artificial intelligence is one of the greatest technological developments of the 21st century and it has the ability to have a positive, but also negative, impact on human life. In this way, the use of artificial intelligence systems can result in the violation of fundamental rights such as equality, privacy, correct process and freedom of expression. In this scenario, the Law needs to respond to the new challenges that arise and offer avant-garde legal responses. One of them is the recognition of human control as a new right to guarantee oversight in the design and development of these systems in order to avoid affecting other existing rights.

KEYWORDS Artificial intelligence, fundamental rights, human control.

Introducción

La inteligencia artificial es considerada uno de los mayores desarrollos tecnológicos del siglo XXI y puede ser entendida como la capacidad de una máquina computa-

cional u ordenador para solucionar problemas complejos y determinados, mediante la implementación de un algoritmo (Almonacid y Coronel, 2019). Enmarcada en lo que se ha catalogado como la Cuarta Revolución Industrial, la inteligencia artificial ocupa la mayoría de los espacios de la vida. A diferencia de los anteriores desarrollos tecnológicos, no solo está dirigida a la industria y la producción masiva de bienes y servicios, sino también a las labores cotidianas, el ocio, las relaciones interpersonales y los espacios más íntimos de la existencia, deshumanizándolos o, por lo menos, tecnificándolos como nunca se había imaginado.

En este escenario, la inteligencia artificial desafía las reglas de la ingeniería, la medicina y las ciencias y también del derecho, a partir de cuestionamientos sobre cómo los ordenamientos jurídicos deben responder a estos nuevos desarrollos. Es ahí, entonces, donde cabe hacer hincapié en la protección de los derechos hasta ahora conocidos y, a su vez, plantear la necesidad de reconocer unos nuevos que permitan proteger los intereses del ser humano, ya no en una realidad análoga sino en escenarios artificiales.

La cuarta generación de derechos es la respuesta que se ha ofrecido a tales interrogantes. La vigencia de los derechos hasta ahora reconocidos y el surgimiento de nuevos derechos en escenarios digitales y de inteligencia artificial son el punto de diferenciación entre la cuarta generación de derechos y las generaciones ya existentes. Lo anterior abre la puerta a un nuevo proceso de reconocimiento de derechos en los escenarios de inteligencia artificial que enmarcan lo aquí propuesto: la necesidad de reconocer el control humano sobre la inteligencia artificial como un nuevo derecho.

El control humano como un nuevo derecho es la respuesta jurídica a la necesidad de mantener una veeduría, inspección o intervención humana en el desarrollo de los sistemas y las tecnologías que se encuadran en la inteligencia artificial, con el fin de garantizar la protección de los bienes jurídicos ya tutelados, como lo son la igualdad, la libertad de expresión, la privacidad, el debido proceso, entre otros. Esto significa que el valor del control humano como derecho radica en su sentido instrumental, o sea, se justifica pues protegería derechos intrínsecos al ser humano, en escenarios aún desregulados jurídicamente, como los de la inteligencia artificial.

Teniendo presente lo anterior, este artículo expone un primer acercamiento teórico y jurídico a lo que significa el control humano como derecho, con el fin de evidenciar cómo la inteligencia artificial representa una amenaza para los derechos del ser humano y, en este sentido, cómo el control humano es una respuesta para mitigar tales amenazas. Para esto, el texto se divide en cuatro apartados. El primero está dirigido a justificar la necesidad de reconocer el control humano como un nuevo derecho. El segundo pretende conceptualizar el control humano a partir de una recopilación de pronunciamientos que al respecto han realizado organizaciones internacionales gubernamentales y no gubernamentales y empresas líderes en los desarrollos de inteligencia artificial. Por su parte, el tercer apartado del texto está orientado a exponer

brevemente los objetivos que perseguiría el control humano como un derecho en el ordenamiento jurídico para, finalmente, terminar con algunas conclusiones.

¿Por qué reconocer el control humano como un nuevo derecho?

La inteligencia artificial es un término frecuentemente nombrado que se ha incorporado a la vida cotidiana con la misma velocidad que sus desarrollos pues sus avances tecnológicos sorprenden cada día a la humanidad. Esta nueva tecnología ha invadido, bien sea de manera positiva o negativa, casi todas las esferas de la vida, tanto pública como privada: los negocios, las relaciones interpersonales, la medicina, la educación y hasta las actividades del hogar, se han visto impactadas por sistemas o dispositivos que cuentan con inteligencia artificial (Penrose, 2019).

Pero, entonces, ¿qué es inteligencia artificial? Si bien es una pregunta aparentemente sencilla, tiene variadas respuestas sobre las que todavía no hay un consenso. Por lo tanto, hasta este punto, se entenderá por inteligencia artificial el conjunto de técnicas, algoritmos y herramientas que permiten resolver problemas para los que, *a priori*, es necesario cierto grado de inteligencia, en el sentido de que son problemas que suponen un desafío, incluso, para el cerebro humano (García, 2012). Sin embargo, más allá de definir la inteligencia artificial, lo que aquí preocupa son los efectos negativos que puede tener sobre los derechos humanos y, por esta razón, conviene preguntarse si es necesario crear nuevos derechos, reglas y protecciones referentes a estos nuevos sistemas artificiales, también.

Sobre la inteligencia artificial existen certezas y, al mismo tiempo, incertidumbres. Hay certeza en tanto ha facilitado la vida humana realizando tareas complejas para el individuo en un tiempo reducido y porque ha contribuido a la resolución de grandes problemas de la humanidad como el empleo y el comercio, entre otros (Oppenheimer, 2018). En otras palabras, ha marcado un hito en la evolución y en los desarrollos tecnológicos más importantes del último siglo. No obstante, también ha dejado en evidencia los riesgos que se corren con su uso; por ejemplo, al afectar seriamente bienes jurídicos tutelados como consecuencia de fallas en los sistemas.

Muestra de lo anterior son algunos casos en donde se presentan sesgos discriminatorios de los datos ingresados a los sistemas de inteligencia artificial, los cuales utilizan *big data* y «toman decisiones» como: asociar los cargos mejor remunerados con hombres y puestos de trabajo peor remunerados con mujeres o considerar como sospechosos de delitos a hombres negros, por encima de los blancos. En situaciones como estas, ¿quién garantiza el derecho a la igualdad y no discriminación de los implicados? o ¿quién evita el procesamiento sesgado de los datos?

Para mayor claridad, Urueña (2019: 16) ilustra la manera como la inteligencia artificial puede reproducir factores de discriminación, vulnerando el derecho a la igualdad:

Piénsese, por ejemplo, en instancias de discriminación algorítmica. El Compas [Correctional Offender Management Profiling for Alternative Sanctions], la plataforma de predicción de reincidencia reseñada antes, señalaba sistemáticamente a los hombres negros como más proclives a la reincidencia. El proceso de reconocimiento de fotos de Google, un ejemplo típico de *machine learning*, categorizó fotos de personas de raza negra como «gorilas». Y el algoritmo de anuncios de Google tendió en un momento dado a mostrar a las mujeres menos anuncios de los trabajos mejor pagos.

Otro ejemplo de lo propuesto es el robot virtual de Microsoft, llamado Tay, que en su cuenta de Twitter, @TayandYou, está programado para almacenar y procesar datos procedentes de sus conversaciones con tuiteros humanos y así perfeccionar su lenguaje, aptitudes y actitudes *millennial* para parecer cada vez más una joven real. Lo que no se esperaba era que Tay fuera capaz de alabar a Hitler, insultar a las feministas con mensajes como «Odio a las feministas, deberían morir todas y pudrirse en el infierno» o «Bush generó el 11S y Hitler habría hecho un trabajo mejor que el mono [Barack Obama] que tenemos ahora» (Malvar, 2017).

Ejemplos como los anteriores evidencian que las decisiones algorítmicas pueden ser sesgadas o discriminatorias debido a precondiciones perjudiciales, diseños inadecuados o conjuntos de datos insuficientes o mal seleccionados (Executive Office of the President, 2016). A medida que los sistemas basados en algoritmos adoptan decisiones automatizadas, el posible sesgo se vuelve masivo y, en muchos casos, pasa desapercibido mientras se vuelve viral. En una sociedad densamente interconectada, la viralidad de la red actúa como un amplificador de los efectos nocivos, estos se expanden y el daño aumenta enormemente, lo cual hace que la reversibilidad de las acciones se vuelva más improbable y menos factible (Rodríguez de las Heras Ballell, 2019).

Los patrones de discriminación estructural existentes se pueden reproducir y agravar en situaciones particulares con estas tecnologías. Por ejemplo, se afirma que los algoritmos inevitablemente toman decisiones sesgadas, en tanto su diseño y funcionalidad reflejan los valores de su diseñador, lo cual lleva a que el desarrollo de la inteligencia artificial no sea un camino neutral y lineal (Mittelstadt y otros, 2016). Igualmente, se afirma que los algoritmos imitan el proceso de toma de decisiones humanas y que están capacitados para aprender de los éxitos pasados, lo cual puede incluir sesgos existentes. En este sentido, la discriminación en la inteligencia artificial puede considerarse una expresión consistente y repetida de una preferencia, creencia o valor particular en la toma de decisiones, lo que, a su vez, ocasiona efectos sociales problemáticos.

No obstante, también existen posiciones menos radicales que reconocen que los sistemas de inteligencia artificial son máquinas sintácticas y no semánticas, lo que significa que pueden realizar acciones y manipulaciones, pero sin atribuirles ningún

significado, pues este es dado por los operadores humanos solamente. Al ser máquinas irreflexivas, no pueden analizar ni comprender las implicaciones éticas de su comportamiento (King y otros, 2020) por lo que por sí mismas no reproducen sesgos discriminatorios, sino que todo dependerá de los datos con los que se nutra el sistema y la preservación de atributos como la objetividad e imparcialidad en su recolección y tratamiento.

Por su parte, surgen preocupaciones sobre la garantía del debido proceso en los procedimientos judiciales y administrativos que utilizan tecnologías, como los sistemas jurídicos expertos,¹ para tomar decisiones tan trascendentales como definir la condena de privación de libertad de una persona o imponer multas y sanciones administrativas a los ciudadanos. Estos sistemas fueron diseñados, entre otras cosas, para apoyar la toma de decisiones de los jueces y emitir sentencias en los diferentes juicios que realizan, a partir de un prototipo de sentencia cuya base de conocimiento está integrada por los requisitos de forma y fondo de una determinada sentencia del derecho.

La pregunta es, entonces, cómo garantizar el debido proceso cuando humanos, y de manera particular juezas y jueces, hacen uso de sistemas de inteligencia artificial, robots o máquinas para apoyarse en la toma de decisiones; considerando que estos sistemas no tienen la capacidad interpretativa de los hechos y las circunstancias que sí tiene un humano. En efecto, estos sistemas solo están diseñados a partir de bases de datos del procedimiento judicial, datos de la demanda, contestación y análisis de las pruebas documentales, confesionales, testimoniales y presuncionales.

En esta misma línea, se destaca que la Fiscalía General de la Nación de Colombia se encuentra implementando la herramienta Prisma (Perfil de Riesgo para Solicitud de Medidas de Aseguramiento), la cual ha sido diseñada para predecir el riesgo de reincidencia criminal de personas capturadas, con el objetivo de reducir errores en las decisiones sobre medidas de aseguramiento privativas de la libertad. Desde un punto de vista de eficiencia, los propósitos de Prisma son reducir la reincidencia criminal y hacer un uso más razonable de los escasos cupos carcelarios y, desde un punto de vista de justicia, hacer un uso más proporcional de las medidas de aseguramiento intramural, intentando que se concentren en aquellos individuos con altos niveles objetivos de riesgo de reincidencia criminal.

1. Los sistemas expertos en la informática jurídica son la estructuración de conocimientos especializados en términos jurídicos que, acoplados a un mecanismo de inferencia, extraen conclusiones de información suministrada de esta área del conocimiento en forma de preguntas y respuestas. Dicho de otro modo, un sistema experto es aquel que, partiendo de ciertas informaciones proporcionadas por un especialista en la materia considerada, pretende resolver problemas que se presentan al interior de un «dominio» específico. Así, se busca que un conjunto de programas sea capaz de alcanzar los resultados de un experto humano para una tarea dada (Batista y otros, 2019: 132).

Ahora bien, con la herramienta Prisma surgen preguntas acerca de cuáles son las «características del individuo» que se tienen en cuenta o si las «capturas previas», sin distinguir si fueron o no legalizadas, deben ser parte de las variables tenidas en cuenta para predecir la reincidencia criminal. Lo anterior puede ser especialmente problemático si se tiene en cuenta que la Corte Constitucional de Colombia ha considerado inconstitucional que en estas decisiones se tengan en cuenta aspectos correspondientes al «derecho penal de autor» en el cual se sanciona a las personas por sus características personales y no por sus actos (Cepeda y Otálora, 2020).

Del mismo modo, es importante referirse sobre PretorIA, un sistema de inteligencia artificial adoptado por la Corte Constitucional de Colombia que busca mejorar su proceso de selección de tutelas. Este asunto, que tradicionalmente demanda un largo trabajo a las personas encargadas de seleccionar los casos que la Corte revisará con el fin de emitir una sentencia, con PretorIA se facilitará dado que este realiza un primer análisis de las sentencias de tutela para entregarle información más procesada a quienes se encargan de identificar cuáles ameritan ser seleccionadas (DeJusticia, 2021).

Al respecto, cabe preguntar cómo este sistema incorpora los criterios que previamente la Corte Constitucional de Colombia definió en su reglamento para seleccionar tutelas, los cuales obedecen a i) criterios objetivos, como la unificación de jurisprudencia o la necesidad de pronunciarse frente a una determinada línea jurisprudencial; ii) criterios subjetivos, como la urgencia de proteger un derecho fundamental y, iii) criterios complementarios, como la lucha contra la corrupción.² Todos estos son criterios que obedecen a la importancia de respetar el debido proceso en todas las decisiones judiciales, incluida la selección de una tutela para revisión por parte de la Corte Constitucional de Colombia.

En suma, surge la necesidad de mantener protegido el derecho al debido proceso cuando se hace uso de sistemas de inteligencia artificial en procesos judiciales y administrativos, lo que abre el debate a cuestiones como la competencia y capacidad del sujeto de decisión, que debería mantenerse en un ser humano y se ha denominado «reserva de humanidad» en el derecho administrativo (Ponce, 2019); el respeto de las garantías procesales en todas las etapas de decisión; y el cumplimiento de requisitos formales en dichos procesos con asistencia de inteligencia artificial.

Por otro lado, surgen preguntas igualmente importantes sobre el destino y el tratamiento de los datos personales entregados a los sistemas artificiales: ¿Quién conoce la vida privada de las personas?, ¿qué pasa con los datos entregados a los robots que realizan tareas en los hogares?, ¿quién controla el traspaso de datos a través del internet de las cosas? Estas preguntas resumen las preocupaciones que existen sobre el derecho a la privacidad, el cual se ve amenazado por la entrega de información

2. «ABECÉ de la acción de tutela», Corte Constitucional de Colombia, 2020, disponible en <https://bit.ly/3xduLoE>.

privada a los sistemas de inteligencia artificial sin tener pleno conocimiento sobre el destinatario final y el uso que se le dará. En este sentido, es claro que se pierde el control sobre la información una vez que se entrega.

Como ejemplo de los riesgos que implica el *big data* para los derechos a la privacidad y a la protección de datos personales «se puede tomar el sonado caso de la cadena de almacenes Target, cuya analítica de datos le permitió inferir, a partir de los datos de compras de sus clientes, el embarazo de una cliente adolescente, incluso antes de que sus padres lo supieran» (Newman y Ángel, 2019). Situaciones como estas son cada vez más comunes, pues constantemente se está entregando información personal a los sistemas artificiales: al crear una cuenta de perfil, al suministrar la ubicación para la entrega de una compra, los datos de pago, las cuentas bancarias, la fecha y hora de acceso a los servicios, las personas con quienes se interactúa, los sitios web más frecuentados, entre otras. En suma, con la inteligencia artificial también surge la preocupación sobre el control que se debe tener de los datos personales y la vigilancia necesaria para asegurar su uso correcto y con sentido.

Por su parte, sobre el derecho a la libertad de expresión es importante señalar que con la creación de nuevas tecnologías han surgido escenarios digitales para su ejercicio, como blogs, tuits, videos, memes o incluso las reacciones en redes sociales (Gutiérrez, 2016). En general, Internet se ha consolidado como un espacio idóneo para la libertad de expresión, en tanto facilita la libertad de acceso, la multiplicidad de formatos de información, la descentralización en la producción y consumo de información, la posibilidad de interacción de los usuarios en tiempo real y la neutralidad de la información (Corte Constitucional de Colombia, 2015).³

No obstante, los espacios digitales también son escenarios para la creación de amenazas y vulneraciones a la libertad de expresión con restricciones a contenidos publicados, intromisiones ilegítimas de terceros, falta de neutralidad en las publicaciones, censura de los contenidos publicados, entre otros. En general, es posible decir que la libertad de expresión se puede ver amenazada por los intereses de quienes dominan los espacios digitales, reconociendo que estos son administrados y coordinados por empresas privadas y, a su vez, por los Estados, quienes mantienen sus potestades y poderes exorbitantes, no solo en espacios físicos sino también artificiales.

Como se señaló, la inteligencia artificial puede ser imperfecta y sus errores pueden tener consecuencias perjudiciales en el ser humano, en especial, pueden vulnerar derechos humanos ya reconocidos y tutelados por los ordenamientos jurídicos. Dicha afectación se debe en gran medida a la ausencia de control y vigilancia humana en el diseño y desarrollo de los sistemas de inteligencia artificial, lo cual los dota de un alto grado de autonomía. En otras palabras, su carácter autónomo puede ser utili-

3. Corte Constitucional de Colombia, sentencia T-277/2015, disponible en <https://bit.ly/3iWPS8x>.

zado por sus fabricantes para limitar la intervención humana y así evitar la afectación de derechos humanos, como ya se ha señalado.

En este escenario, el control humano es la respuesta que se propone para solucionar el problema jurídico que representa la amenaza y vulneración de bienes y derechos tutelados. Actualmente, el reconocimiento del control humano como derecho y principio de la inteligencia artificial es un tema relevante de discusión en los países con altos desarrollos al respecto y en las empresas líderes, pues se identifica la necesidad de que los actores de inteligencia artificial respeten el Estado de derecho, los derechos humanos y los valores democráticos, durante todo su ciclo de vida (OCDE, 2020). Empresas como IBM, Microsoft, Telia Company e IA Latam se han sumado a esta discusión junto con países como Canadá, Japón, Estados Unidos y la comunidad política de la Unión Europea, a partir de declaraciones públicas y privadas en las que han manifestado la necesidad de considerar el control humano en la formulación y desarrollo de sistemas de inteligencia artificial.⁴

Teniendo presente lo anterior, es preciso preguntarse: ¿por qué catalogar al control humano como un derecho? O, más bien, ¿por qué no basta con las declaraciones ya existentes al respecto? Las respuestas a estas interrogantes son, en esencia, la justificación de su reconocimiento. Como ya se señaló, la inteligencia artificial representa una amenaza a bienes jurídicos tutelados, lo que hace necesaria la intervención jurídica. Esto último quiere decir que no es posible dejar a voluntad de las empresas —quienes en definitiva persiguen intereses particulares— la decisión de privilegiar los derechos humanos sobre otros intereses y mucho menos permitir que la garantía de estos dependa de la empresa con la que se contratan los sistemas de inteligencia artificial. En

4. Véase: «IBM Everyday ethics for artificial intelligence», 2019, disponible en <https://ibm.co/3qxJCdT>; «Guiding principles on trusted ai ethics», Telia Company, 2019, disponible en <https://bit.ly/3wIB7oP>; «Responsible AI principles from Microsoft», 2018, disponible en <https://bit.ly/3oh4ziq>; «Declaración de ética IA-Latam para el diseño, desarrollo y uso de la inteligencia artificial», disponible en <https://bit.ly/3naDBC3>; «Preparing for the future of artificial intelligence», Executive Office of the President National Science and Technology Council Committee on Technology, 2016, disponible en <https://bit.ly/3n8dCes>; «Human Rights in the age of artificial intelligence», Access Now, 2018, disponible en <https://bit.ly/3C5rpXq>; «The Toronto declaration: Protecting the right to equality and non-discrimination in machine learning systems», Amnesty International y Access Now, 2018, disponible en <https://bit.ly/3wOXceg>; «Social principles of human-centric AI», Japanese Cabinet Office, Council for Science, technology and innovation, 2019, disponible en <https://bit.ly/3wICMDB>; «European ethical charter on the use of artificial intelligence in judicial systems and their environment Council of Europe», European Commission for the Efficiency of Justice (CEPEJ), 2018, disponible en <https://bit.ly/2YHSpP9>; «Communication from the commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the regions: Artificial intelligence for Europe», 2018, disponible en <https://bit.ly/3Ce8wBD>; «Directrices éticas para una IA fiable», Grupo de expertos de alto nivel sobre inteligencia artificial, Comisión Europea, 2018, disponible en <https://bit.ly/3kxlYux>.

este mismo sentido, es necesaria una directriz general internacional o, por lo menos, estatal, que cobije a todas las empresas y desarrollos de inteligencia artificial.

Dicho de otro modo, los Estados tienen el deber de garantizar el goce efectivo de los derechos humanos. Así lo señala el Sistema Interamericano de Derechos Humanos: los Estados parte se comprometen a respetar los derechos y libertades de las personas y garantizar su libre y pleno ejercicio. Lo anterior, sin discriminación alguna por motivos de raza, color, sexo, idioma, religión, opiniones políticas o de cualquier otra índole, origen nacional o social, posición económica, nacimiento o cualquier otra condición social (Comisión Interamericana de Derechos Humanos, 1969).

En esta línea, una norma de orden público debe considerar el control humano como un derecho, con sus características propias, es decir, universal, inalienable, irrenunciable, imprescriptible e indivisible. Asimismo, el Estado tiene el deber de intervenir en estos nuevos escenarios artificiales y adoptar las medidas que al respecto ya se han desarrollado en otros Estados con mayores avances en el tema, como los ya citados, y lo dicho por las organizaciones de derecho público internacional.

En suma, la justificación de reconocer el control humano como un derecho radica en el deber que tienen los Estados de proteger los derechos humanos en los sistemas de inteligencia artificial. Esto a partir de leyes, normas de orden general, actos administrativos o pronunciamientos judiciales que garanticen la tutela de los derechos ya reconocidos en estos nuevos escenarios. De esta manera, aunque son necesarios, resultan insuficientes los postulados promulgados por las empresas tecnológicas sobre el control humano como un principio de la inteligencia artificial, pues al estar directamente implicadas en el desarrollo de estos sistemas y conocer de primera mano los efectos adversos que tienen en los derechos humanos, carecen de la legitimidad necesaria para considerar como derecho el control humano.

En este punto, es importante señalar que reconocer el control humano como un derecho lo dota de un carácter general, aplicable a cualquier circunstancia o sistema de inteligencia artificial existente y, además, deja abierta la posibilidad de que se aplique en desarrollos futuros que hoy no son previstos, pero pueden existir en poco tiempo. Es aquí cuando cobra relevancia la necesidad de no solo reconocer el derecho, sino también de regularlo, pues este aspecto implica que se consideren elementos fundamentales —como la naturaleza del derecho de la que se desprende la forma como este debe ser legislado en el ordenamiento jurídico— y las acciones judiciales y administrativas con las que contaría el titular para exigir su protección.

¿Qué es el control humano?

Lo primero que se debe señalar sobre el control humano es que es un concepto en construcción. Las aproximaciones teóricas que se han hecho provienen de empresas,

gobiernos y organizaciones civiles e internacionales⁵ que han reconocido la necesidad de controlar los sistemas de inteligencia artificial. En este sentido, se ha entendido por control humano la capacidad de que intervengan seres humanos durante el ciclo de diseño del sistema de inteligencia artificial y en el monitoreo de su funcionamiento, con el fin de evitar un impacto negativo en los derechos humanos y facilitar el cumplimiento de objetivos como seguridad, protección, transparencia, explicabilidad, justicia, no discriminación y, en general, la promoción de valores humanos dentro de los sistemas de inteligencia artificial (The Public Voice Coalition, 2018).

El control humano comprende la supervisión, participación, revisión y determinación humana. Esto es que los sistemas permanezcan siempre bajo el control humano —incluso de manera *ex post* con la revisión de las decisiones que dichos sistemas determinen— y siempre impulsados por consideraciones basadas en valores (The Public Voice Coalition, 2018). Al respecto, organizaciones como el Observatorio Iberoamericano de Protección de Datos han reconocido la necesidad de prohibir que cualquier decisión que afecte a una persona se adopte exclusivamente por sistemas automatizados, sin intervención de otra persona humana con potestad decisoria para cada caso concreto (Llinás, 2008).

Por su parte, IBM (2019) considera que el juicio humano juega un papel fundamental en un sistema aparentemente objetivo de decisiones lógicas. Así, son los humanos quienes escriben algoritmos, definen el éxito o el fracaso de un sistema, toman decisiones sobre los usos de dicho sistema y quienes, en definitiva, se pueden ver afectados por sus resultados. Cada persona involucrada en la creación de inteligencia artificial, al igual que las empresas que invierten en su desarrollo, es responsable de considerar el impacto de dicho sistema en el mundo, en cualquier caso. Para esto, los diseñadores y desarrolladores de la inteligencia artificial deben implementar mecanismos y salvaguardas, como la capacidad de determinación humana, que sean apropiadas al contexto y consistentes con el estado del arte (OCDE, 2020).

Por otro lado, el G20 (2019) plantea la determinación humana en términos de derechos, es decir, afirma que los individuos tienen derecho a una determinación final hecha por una persona y para esto es necesario que los actores de inteligencia artificial implementen mecanismos que sean apropiados al contexto. En este caso, los humanos deben elegir si quieren delegar la toma de decisiones a los sistemas de inteligencia artificial y cómo desean hacerlo. En suma, hablar de control humano implica considerar un monitoreo humano constante sobre el conjunto de datos y algoritmos que son insumos para el sistema de inteligencia artificial y su desarrollo (Telia Company, 2019).

5. Dentro de este conjunto de organizaciones se encuentran: Organización para la Cooperación y el Desarrollo Económicos, UNI Global Union, European Commission's High-Level Expert Group on Artificial Intelligence, Council of Europe, European Commission for the Efficiency of Justice, The Public Voice Coalition, Telefónica (Raso y otros, 2018).

En este escenario, surgen dos preguntas para comprender el alcance del control humano como un derecho. La primera, ¿por qué es necesario el control humano en los sistemas de inteligencia artificial? Y la segunda, ¿para qué se necesita el control humano en la inteligencia artificial? Como respuesta a la primera pregunta, es preciso señalar que los sistemas de inteligencia artificial son máquinas sintácticas y no semánticas, lo que significa que pueden realizar acciones y manipulaciones, pero sin atribuirles ningún significado. Esto último es realizado solamente por los operadores humanos. Al ser máquinas irreflexivas, no pueden analizar ni comprender las implicaciones éticas de su comportamiento (King y otros, 2020).

Lo anterior se debe a que el ser humano solo interviene los sistemas artificiales en el momento de su diseño, con la programación de los algoritmos que determinan su funcionamiento, pero una vez que entra en operación, el individuo se distancia. Esto se justifica por la naturaleza misma de la inteligencia artificial, pues como ya se anticipó, tiene la capacidad de aprender de forma autónoma, sin intervención humana. En este sentido, la respuesta a la primera pregunta radica en la naturaleza de los sistemas que son autónomos y pueden funcionar sin intervención humana, toda vez que su sistema ya ha sido programado a través de algoritmos, pero como máquinas sintácticas y no semánticas.

Ahora bien, aquí cobra relevancia la segunda pregunta sobre el *para qué* del control humano. Al respecto, Telefónica (2018) ha señalado que los sistemas de inteligencia artificial tienen que permanecer siempre bajo el control humano para que no generen ningún impacto negativo en los derechos humanos. A su vez, IBM (2019) considera que la intervención humana es necesaria para minimizar sesgos algorítmicos: decisiones automatizadas que reproduzcan patrones de discriminación de una sociedad. Por su parte, el G20 (2019) señala que el control humano garantiza que la inteligencia artificial se centre en el ser humano y aporte a una sociedad inclusiva, segura, confiable e innovadora, aun en escenarios artificiales.

Del control humano también se desprenden otros beneficios como diseñar sistemas de inteligencia artificial que cumplan las leyes existentes, controlen los datos ingresados y la privacidad de sus titulares y expliquen los procedimientos usados para la construcción de dichos sistemas.⁶ Esto último se relaciona con la necesidad de informar al usuario, en un lenguaje claro y comprensible, las soluciones ofrecidas por las herramientas de inteligencia artificial para cualquier proceso realizado y para tener derecho de objetarlo (Comisión Europea para la Eficiencia y la Justicia, 2018). Esta claridad y una mayor comprensión digital ayudarán al público a experimentar las ventajas de la inteligencia artificial, así como a optar por no utilizar dichos productos si tienen dudas. Lo anterior remite a un principio de transparencia, a partir

6. UNI Global Union, «Top 10 principles for ethical artificial intelligence», disponible en <https://bit.ly/3csUgtn>.

del cual todas las personas tienen derecho a conocer la base de una decisión de inteligencia artificial que les concierna; lo que incluye el acceso a los factores, la lógica y las técnicas que dieron lugar al resultado.

En general, es posible decir que los sistemas de inteligencia artificial pueden favorecer como obstaculizar los derechos. En situaciones en las que existan riesgos deberá implementarse el control humano para identificar, prevenir, mitigar y explicar cómo se abordan los impactos sobre los derechos humanos y cómo se manejan dichos riesgos. En síntesis, el control humano es un presupuesto para respetar los derechos humanos y los valores democráticos durante todo el ciclo de vida del sistema de inteligencia artificial. Dentro de dichos derechos y valores, internacionalmente reconocidos, se incluyen la libertad, dignidad, autonomía, privacidad y protección de datos, no discriminación e igualdad, diversidad, justicia, justicia social y derechos laborales.

De este modo, el control humano debe ser el mecanismo veedor de la alineación de valores en los sistemas que son altamente autónomos, de modo que siempre sean congruentes con los principios jurídicos y compatibles con la dignidad humana. Así las cosas, la superinteligencia solo debe desarrollarse al servicio de ideales éticos ampliamente compartidos que propendan al beneficio de toda la humanidad (Future of Life Institute, 2017) y, adicionalmente, se deben contrarrestar las aplicaciones potencialmente dañinas o abusivas (IA Latam, 2019).

En síntesis, son múltiples los actores que han manifestado su compromiso con la protección de los derechos humanos en los desarrollos de inteligencia artificial. Estos derechos giran alrededor de los principios de privacidad, responsabilidad, protección y seguridad, transparencia y explicabilidad, equidad y no discriminación, control humano de la tecnología, responsabilidad profesional y promoción de valores humanos (Fjeld y otros, 2020). En este sentido, es posible decir que sus declaraciones han puesto sobre la mesa la necesidad de reconocer nuevos derechos humanos, a partir de la amenaza que puede representar la inteligencia artificial para los ya existentes.

Objetivos del derecho al control humano

Justificar el control humano como un derecho implica identificar los objetivos que debe perseguir; de ello se deriva su categorización en el ordenamiento jurídico y la definición de las obligaciones y garantías que se desprenden. En consonancia, algunos autores señalan que para justificar el control humano como derecho se necesita evidenciar que los objetivos perseguidos son valiosos (Beitz, 2012: 15). Así, se puede establecer que los objetivos que persigue el control humano como derecho son: i) garantizar en los sistemas de inteligencia artificial la tutela de los derechos humanos ya reconocidos; ii) asegurar la veeduría e intervención humana en el diseño y desarrollo de los sistemas de inteligencia artificial y, finalmente, iii) corregir cualquier amenaza o vulneración de los derechos humanos en los sistemas de inteligencia artificial. Estos objetivos se describirán a continuación.

Garantizar la tutela de los derechos humanos ya reconocidos en los sistemas de inteligencia artificial

Como se ha mostrado, los sistemas de inteligencia artificial son tan imperfectos como el ser humano y es que, en tanto son una creación suya, representan y reproducen los mismos sesgos, amenazas o vulneraciones sobre los derechos humanos que se pueden atribuir a las personas. La libertad, la igualdad, la privacidad y el debido proceso son derechos que constantemente están en riesgo en el diseño y desarrollo de los sistemas de inteligencia artificial, bien sea porque son programados con sesgos y datos contradictorios de los derechos humanos o porque, en virtud de su autonomía y sus capacidades de aprendizaje autónomo, absorben de la interacción con el mundo tanto los elementos positivos como los negativos, sin la existencia de filtros o criterios claros de discernimiento.

Por esta razón, surge la necesidad de mantener un control, vigilancia o, si se quiere, inspección humana en el diseño y desarrollo de la inteligencia artificial. No para limitar su autonomía, lo cual iría en contravía de su propia naturaleza, sino para evitar acciones u omisiones que vulneren o amenacen los derechos humanos que ya han sido reconocidos en el ordenamiento jurídico.

De esta manera, la justificación del control humano como derecho no es autónoma, por el contrario, tiene un carácter instrumental pues su existencia solo tiene un sentido jurídicamente válido gracias a su conexión con otros derechos humanos, muchos de ellos fundamentales, y con el propósito de asegurar su tutela en sistemas de inteligencia artificial. Así, lo que se busca con el control humano es extender las garantías, los recursos y la vigilancia de los derechos que hasta el momento tienen los seres humanos en el mundo análogo, a un mundo, si se quiere decir, artificial. Cabe recordar que por ahora este mundo artificial es regido por los intereses privados de las empresas y por las relaciones particulares entre productores, distribuidores y consumidores, en las que la intervención estatal es mínima o hasta inexistente.

Es importante decir que el control humano surge como una respuesta jurídica a una necesidad pública y, al mismo tiempo, a una obligación estatal. Igualmente, es una necesidad pública la intervención de las autoridades para preservar los derechos humanos en escenarios que, hasta ahora, han sido dominados por agentes no estatales como las empresas, por ejemplo. Esto quiere decir que es necesario que las personas tengan garantizados y protegidos sus intereses y derechos humanos en los desarrollos de inteligencia artificial existentes y futuros a través de una cláusula de protección genérica y maleable a las circunstancias y progresos tecnológicos. El derecho al control humano da, entonces, la posibilidad de adaptación a situaciones previstas, pero también a las imprevisibles debido a la velocidad con la que evolucionan los sistemas de inteligencia artificial.

Finalmente, el sentido del control humano como derecho responde a una obli-

gación estatal que se encuentra, para el caso colombiano, en el artículo segundo de la Constitución Política. En virtud de este, es una obligación del Estado «garantizar la efectividad de los principios, derechos y deberes consagrados en la Constitución» (Constitución Política de Colombia, 1991). En otras palabras, es obligación del Estado asegurar la protección de los derechos en todas las circunstancias o los escenarios que, como es lógico, van evolucionando con el tiempo. De este modo, el deber de protección estatal y de intervención mediante instituciones competentes en escenarios de inteligencia artificial, cobra cada día más relevancia y se hace más urgente dar una respuesta integral por parte del Estado.

Asegurar la veeduría e intervención humana en el diseño y desarrollo de los sistemas de inteligencia artificial

Considerar el control humano como un derecho supone la potestad de exigir la intervención de una persona en todas las fases del proceso de creación y desarrollo de los sistemas de inteligencia artificial, con el propósito de evitar vulneraciones o amenazas a los derechos humanos. El objetivo de asegurar la veeduría e intervención humana cumple una función sobre todo preventiva, es decir, corregir acciones u omisiones de los sistemas de inteligencia artificial, antes de que se produzca un daño irremediable que represente una vulneración de derechos.

Lo anterior significa que la intervención humana debe evitar que se materialicen riesgos ya evidenciados que atenten contra los derechos humanos. Este objetivo supone, entonces, dos funciones. En primera instancia, tener la capacidad de advertir riesgos y amenazas a derechos humanos a partir de criterios objetivos y de los desarrollos jurídicos existentes en la normatividad y jurisprudencia colombiana. En segundo lugar, corregir o, si se quiere, remediar anticipadamente la conducta desviada con el propósito de que se ajuste al ordenamiento jurídico y se asegure su correspondencia con los derechos humanos existentes.

Por último, asegurar la veeduría y la intervención humana supone, además de lo anterior, garantizar la presencia y el control de autoridades públicas y privadas, externas a los desarrolladores de sistemas de inteligencia artificial en sus procesos de diseño y desarrollo, con el ánimo de asegurar la defensa de intereses diferentes a los privados. Esto significa que la veeduría debe estar a cargo de actores estatales, la sociedad civil y otras empresas pares que evalúen el respeto a los derechos humanos en los sistemas de inteligencia artificial, para que estos prevalezcan sobre los intereses particulares y lucrativos de las empresas. Para todos los casos, los responsables deben ser sujetos especializados, idóneos e imparciales.

Corregir cualquier amenaza o vulneración de los derechos humanos en los sistemas de inteligencia artificial

A diferencia del objetivo anterior, lo que aquí se pretende con el control humano es que las personas que lo ejerzan puedan corregir vulneraciones que los sistemas de inteligencia artificial hayan producido. Esto pretende evidenciar que el control humano no solo es una tarea preventiva, sino también correctiva. De este modo, es claro que el control humano además debe procurar el restablecimiento de los derechos humanos cuando han sido vulnerados, o sea, debe devolver a la posición inicial a las personas que, por cuenta de los sistemas de inteligencia artificial, han visto menoscabados sus derechos e intereses jurídicamente tutelados.

En este punto, se advierte que para el cumplimiento de este objetivo se deben tener consideraciones fundamentales como: i) el momento en el que se determina la vulneración del derecho y su relación con la cadena de fabricación, distribución y uso de los productos o sistemas; ii) la competencia que debe ostentar una persona para decidir la manera y el motivo por el que se ha vulnerado un derecho humano, sin que esto vaya en detrimento del debido proceso y el derecho a defensa del supuesto vulnerador y, finalmente, iii) la determinación de las medidas que se deben implementar para restablecer la posición inicial de la persona a la que se le han vulnerado sus derechos por la inteligencia artificial.

Para terminar, es preciso señalar que el control humano es un derecho que se invoca ante las autoridades, tanto administrativas como judiciales, para que en virtud de sus competencias, determinen la forma como se deben corregir las vulneraciones o amenazas a derechos. En otras palabras, esto significa que el control humano es solo el instrumento para que las autoridades decidan las medidas correctivas o, si es el caso, sancionatorias, cuando no se ha respetado el derecho al control humano en los sistemas de inteligencia artificial.

Conclusiones

De lo aquí expuesto, se concluye que el control humano se puede considerar como una respuesta jurídica vanguardista a la necesidad de garantizar la tutela de los derechos e intereses jurídicamente tutelados en escenarios de inteligencia artificial. La autonomía y la capacidad de aprendizaje de dichos sistemas hacen inminente la necesidad de mantener un control humano en los diferentes procesos de creación y desarrollo de la inteligencia artificial con el objetivo de evitar que afecte derechos humanos o que, en caso de tal afectación, se pueda remediar.

Sobre el control humano, hoy se concluye que se tienen más preguntas que respuestas. Como ya se señaló, este es un concepto teórico y jurídico que aún está en construcción y sobre el que poco se ha hablado. Por esta razón, es necesaria una

discusión oportuna capaz de mantener la velocidad de los avances tecnológicos de la inteligencia artificial. Este caso, como muchos otros, es evidencia de la importancia de que el derecho responda a las necesidades y evoluciones sociales y tecnológicas, antes de que se produzcan daños irremediabiles y vulneraciones a derechos humanos.

Igualmente, es necesario delimitar las implicaciones jurídicas que supone el reconocimiento del control humano como un derecho, lo que se traduce en la definición del titular de tal derecho, quien puede exigir garantías y prerrogativas a las autoridades, invocándolo. Así, con el control humano se lograría la protección de derechos que se pueden ver amenazados con el uso de la inteligencia artificial.

Correlativamente, en virtud del control humano, surgen deberes para el Estado y los particulares, con el propósito de garantizar el nuevo derecho en escenarios de inteligencia artificial. Esto supone la definición por parte del Estado de las responsabilidades que se deben asumir para garantizar el control humano. Asimismo, es necesaria una reestructuración institucional al interior de los Estados para delimitar las autoridades que asumirían la inspección, vigilancia y, si es necesaria, la sanción sobre los agentes públicos y privados que participan en la cadena de producción y desarrollo de la inteligencia artificial.

En suma, el reconocimiento del control humano como derecho en la inteligencia artificial es un camino poco explorado del que se derivan ajustes jurídicos e institucionales que comprometen al Estado y a agentes privados. En este sentido, propiciar la discusión académica y jurídica alrededor de su reconocimiento, es el primer paso para que los Estados extiendan la protección de los derechos humanos en escenarios de inteligencia artificial, para lo que el control humano se presenta como una nueva respuesta jurídica.

Referencias

- ALMONACID, Juan Jorge y Yeisson Coronel (2019). «Aplicabilidad de la inteligencia artificial y la tecnología *blockchain* en el derecho contractual privado». *Revista de Derecho Privado*, 38: 119-142. DOI: [10.18601/01234366.n38.05](https://doi.org/10.18601/01234366.n38.05).
- BATISTA HERNÁNDEZ, Noel, Corina Elena Navarrete Luque, Carmen Magaly León Segura, Manuel de Jesús Real, José Antonio Chiriboga Hungría y Jesús Estupiñán Ricardo (2019). «La toma de decisiones en la informática jurídica basado en el uso de los Sistemas Expertos». *Revista Investigación Operacional*, 40 (1): 131-139. Disponible en <https://bit.ly/3CcPKe3>.
- BEITZ, Charles (2012). *La idea de los derechos humanos*. Barcelona: Marcial Pons.
- CEPEDA, Manuel José y Guillermo Otálora (2020). «Modernización de la administración de justicia través de la inteligencia artificial». Fedesarrollo: Centro de Investigación Económica y Social. Disponible en <https://bit.ly/32XZMBh>.
- DEJUSTICIA (2021). «Conoce nuestra investigación sobre PretorIA, la tecnología

- que incorpora la inteligencia artificial a la Corte Constitucional». Disponible en <https://bit.ly/3ycSkQo>.
- EXECUTIVE OFFICE OF THE PRESIDENT (2016). «Big Data: A report on algorithmic systems, opportunity, and civil rights». Disponible en <https://bit.ly/3icsMNp>.
- FJELD, Jessica, Nele Achten, Hannah Hilligoss, Adam Nagy y Madhulika Srikumar (2020). «Principled artificial intelligence: Mapping consensus in ethical and rights-based approaches to principles for AI». Berkman Klein Center for Internet & Society. Disponible en <https://bit.ly/2YCYot5>.
- FUTURE OF LIFE INSTITUTE (2017). «Asilomar AI principles». Disponible en <https://bit.ly/33SKw7Y>.
- G20 (2019). «G20 ministerial statement on trade and digital economy». Disponible en <https://bit.ly/33XovEX>.
- GARCÍA, Alberto (2012). *Inteligencia artificial: Fundamentos, prácticas y aplicaciones*. Madrid: RC Libros.
- GUTIÉRREZ, Paulina del Pilar (2016). «Desafíos que enfrenta la libertad de expresión e información en el entorno digital». *Dfensor*, 14 (6): 37-41.
- KING, Thomas C., Nikita Aggarwal, Mariarosaria Taddeo y Luciano Floridi (2020). «Artificial intelligence crime: An interdisciplinary analysis of foreseeable threats and solutions». *Science and Engineering Ethics*, (26): 89-120. DOI: [10.1007/s11948-018-00081-0](https://doi.org/10.1007/s11948-018-00081-0).
- MALVAR, Aníbal (2017). «¿Qué fue de Tay, la robot de Microsoft que se volvió nazi y machista?». *Público*, 12 de agosto. Disponible en <https://bit.ly/3kY9U3h>.
- MITTELSTADT, Brent Daniel, Patrick Allo, Mariarosaria Taddeo, Sandra Wachter y Luciano Floridi (2016). «The ethics of algorithms: Mapping the debate». *Big data & Society*, 3 (2): 1-21. DOI: [10.1177/2053951716679679](https://doi.org/10.1177/2053951716679679).
- NEWMAN, Vivian y María Paula Ángel (2019). *Rendición de cuentas de Google y otros negocios en Colombia: La protección de datos personales en la era digital*. Documentos 48. Bogotá: Dejusticia.
- LLINÁS, Emilio (2008). «Declaración de derechos del ciberespacio». Observatorio Iberoamericano de Protección de Datos. Disponible en <https://bit.ly/33WGXxE>.
- OCDE, Organización para la Cooperación y el Desarrollo Económicos (2020). «Recommendation of the council on artificial intelligence: OECD Legal instruments». Disponible en <https://bit.ly/3j34yTw>.
- OPPENHEIMER, Andrés (2018). *¡Sálvese quien pueda!* Bogotá: Nomos.
- PENROSE, Roger (2019). *La nueva mente del emperador*. Madrid: Debolsillo.
- PONCE, Juli (2019). «Inteligencia artificial, derecho administrativo y reserva de humanidad: Algoritmos y procedimiento administrativo debido tecnológico». *Revista General de Derecho Administrativo*, 50. Disponible en <https://bit.ly/2YK54B6>.
- RASO, Filippo A., Hannah Hilligoss, Vivek Krishnamurthy, Christopher Bavitz y Le-

- vin Kim (2018). «Artificial intelligence & human rights: Opportunities & risks». *Berkman Klein Center Research Publication*, 6. DOI: [10.2139/ssrn.3259344](https://doi.org/10.2139/ssrn.3259344).
- RODRÍGUEZ DE LAS HERAS BALLELL, Teresa. (2019). «Legal challenges of artificial intelligence: Modelling the disruptive features of emerging technologies and assessing their possible legal impact». *Uniform Law Review*, 24 (2): 302-314. DOI: [10.1093/ulr/unzo18](https://doi.org/10.1093/ulr/unzo18).
- THE PUBLIC VOICE COALITION (2018). «Universal guidelines for artificial intelligence». Disponible en <https://bit.ly/3cyotEG>.
- URUEÑA, René (2019). «Autoridad algorítmica: ¿Cómo empezar a pensar la protección de los derechos humanos en la era del *big data*?». *Latin American Law Review* 2: 99-124. DOI: [10.29263/lar02.2019.05](https://doi.org/10.29263/lar02.2019.05).

Sobre la autora

CAROLINA SÁNCHEZ VÁSQUEZ es abogada titulada por la Universidad Escuela de Administración, Finanzas e Instituto Tecnológico (EAFIT), Colombia, con énfasis en Derecho Público. Ha sido miembro del Observatorio Parlamentario Antioquia Visible y forma parte del Centro de Análisis Político de la Escuela de Humanidades de la Universidad EAFIT. Su correo electrónico es csanch41@eafit.edu.co.  <https://orcid.org/0000-0002-3896-1857>.

JOSÉ ALBERTO TORO-VALENCIA es abogado. Doctor en Derecho por la Universidad Pompeu Fabra, España. Profesor asociado de Derecho Internacional, Universidad EAFIT, Colombia. Su correo electrónico es jatoro@eafit.edu.co.  <https://orcid.org/0000-0001-6596-9726>.

La *Revista Chilena de Derecho y Tecnología* es una publicación académica semestral del Centro de Estudios en Derecho Informático de la Facultad de Derecho de la Universidad de Chile, que tiene por objeto difundir en la comunidad jurídica los elementos necesarios para analizar y comprender los alcances y efectos que el desarrollo tecnológico y cultural han producido en la sociedad, especialmente su impacto en la ciencia jurídica.

EDITOR GENERAL

Daniel Álvarez Valenzuela
(dalvarez@derecho.uchile.cl)

SITIO WEB

rchdt.uchile.cl

CORREO ELECTRÓNICO

rchdt@derecho.uchile.cl

LICENCIA DE ESTE ARTÍCULO

Creative Commons Atribución Compartir Igual 4.0 Internacional



La edición de textos, el diseño editorial
y la conversión a formatos electrónicos de este artículo
estuvieron a cargo de Tipografía
(www.tipografica.io).